

FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Predicting customer response to cross-market discounts using ensemble methods

Filipe Carlos Neves Mota



Mestrado Integrado em Engenharia Informática e Computação

Orientador: Vera Lucia Miguéis Oliveira e Silva

Junho de 2015

© Filipe Mota, 2015

Predicting customer response to cross-market discounts using ensemble methods

Filipe Carlos Neves Mota

Mestrado Integrado em Engenharia Informática e Computação

Aprovado em provas públicas pelo Júri:

Presidente: Rui Camacho

Vogal Externo: Pedro Abreu

Orientador: Vera Lucia Miguéis Oliveira e Silva

29 de Junho de 2015

Resumo

A globalização dos mercados e o crescente número de empresas tem tornado a área do retalho extremamente competitiva. Esta realidade é transversal aos vários setores de retalho, como são exemplo o sector do vestuário, eletrodomésticos, alimentação, etc.

Para se manterem competitivas, as empresas de retalho têm procurado novas formas de captar a atenção do público. Diferentes técnicas publicitárias e campanhas promocionais são exemplos de estratégias utilizadas para atrair novos clientes e para preservar os atuais. Uma das novas técnicas utilizadas consiste no recurso aos denominados *Cross-Market Discounts*. Estes envolvem duas empresas, com um público-alvo semelhante, que oferecem descontos interligados, isto é, ao efetuar uma transação na primeira empresa esta emitirá um cupão de desconto para ser utilizado numa transação na segunda empresa.

O objetivo desta tese é construir modelos que permitam à empresa em estudo prever a resposta dos seus clientes a este tipo de campanhas. A empresa em estudo trata-se de uma empresa de retalho alimentar e que tem uma parceria com uma gasolinheira. Os modelos serão suportados por diversas técnicas de *data mining*, e por forma a melhorar a performance destas técnicas, serão implementados métodos *ensemble*.

O contributo desta tese passa pela implementação de métodos *ensemble* por forma a melhorar os modelos de previsão da adesão aos *Cross-Market Discount*.

Abstract

The globalization of markets and the growing number of companies increased the competition among retail companies. This reality affects all sectores of retail, from clothing to grocery.

New ways to keep and gain customers help companies keep competitive. Diferent strategies like publicity and discounts are examples of strategies used by companies. One of the most recent strategies is based on cross-market discounts. This strategy consists of offering linked discounts in unrelated markets that have the same target customers but are not in direct competition with each other.

The purpose of this thesis is to build models that allow the company to predict the response of customers to the cross-market discounts. The company in study is a grocery retailer with a partnership with a gas company. The models will be supported by several data mining techniques and to enhance the performance, ensemble methods will be used.

The contribution of this thesis is the implementation of ensemble methods in order to improve the models that predict the response to the cross-market discounts.

Agradecimentos

Gostaria de agradecer, em primeiro lugar, à professora Vera Miguéis por supervisionar o trabalho realizado nesta tese, por todos os conselhos e conhecimento partilhados e pela sua constante disponibilidade. Gostaria também de agradecer a todos os professores, que encontrei ao longo do meu percurso académico, pelo conhecimento que me transmitiram e por estarem sempre disponíveis para ajudar.

Gostaria também de agradecer a todos os meus colegas, que ao longo do meu percurso académico, na faculdade, me acompanhar e ajudaram, em particular ao Ruben Aguiar, Vitor Santos, Pedro Simões e Diogo Basto.

Gostaria ainda de agradecer à minha família, em especial aos meus pais e ao meu irmão. Por fim gostaria de agradecer a todos os meus amigos pelo apoio e por todos os momentos de lazer e diversão que providenciaram ao longo destes anos. Por fim, gostaria de dedicar um agraciamento especial à Bruna por todo o apoio, incentivo, motivação e por todos os momentos que passamos juntos.

Conteúdo

Introdução.....	1
1.1 Contexto/Enquadramento.....	1
1.2 Motivação e Objetivos	2
1.3 Contribuições	3
1.4 Estrutura da Dissertação	3
Enquadramento Teórico.....	5
2.1 <i>Customer Relationship Management</i>	5
2.2 Extração de Conhecimento.....	7
2.3 <i>Data Mining</i>	8
2.4 Classificação	9
2.4.1 Regressão Logística.....	10
2.4.2 Árvores de Decisão	10
2.4.3 <i>Support Vector Machines</i>	11
2.4.4 Redes Neurais.....	12
2.5 Métodos <i>Ensemble</i>	14
2.5.1 <i>Random Forests</i>	15
2.5.2 <i>Bagging</i>	16
2.5.3 <i>Boosting</i>	16
2.5.4 <i>Hybrid Ensemble</i>	16
2.6 Conclusão	17
Caso de Estudo	19
3.1 Introdução	19
3.2 Descrição da empresa.....	19
3.3 Tratamento dos Dados.....	20
3.4 Análise Preliminar de Comportamentos.....	22
3.5 Metodologia	25
3.6 Resultados	26
3.7 Discussão.....	27
3.8 Conclusão	28

Conclusões e Trabalho Futuro	29
4.1 Satisfação dos Objectivos.....	29
4.2 Limitações / Trabalho Futuro	30
Referências.....	31

Lista de Figuras

Figura 1 - Etapas do processo de extração de conhecimento de base de dados (Fayyad et al., 1996)	8
Figura 2 - Árvore de decisão (Han and Kamber, 2006)	11
Figura 3 - Exemplo de um problema de duas dimensões (Cortes e Vapnik, 1995)	12
Figura 4 - Arquitetura de um neurónio	12
Figura 5 - Funções de ativação (Negnevitsky, 2004)	13
Figura 6 - Exemplo de <i>Single layer feed-forward network</i>	13
Figura 7 - Exemplo de <i>Multiple layer feed-forward network</i>	13
Figura 8 – Exemplo de <i>Competitive or recurrent networks</i>	14
Figura 9 - <i>Ensemble</i> de redes neuronais (Maclin e Opitz, 1999)	15
Figura 11 - Gráfico de barras relativo à classificação dos clientes	23
Figura 10 - Histogramas do Número Médio de Transações p/Mês	23
Figura 12 - Histogramas do Valor Médio p/Transação	24

Lista de Tabelas

Tabela 1 - Valor médio das variáveis para clientes aderentes e não aderentes	24
Tabela 2 - AUC de cada técnica implementada	27

Abreviaturas e Símbolos

AD	Árvores de Decisão
AUC	<i>Area Under Curve</i>
CRM	<i>Customer Relationship Management</i>
RL	Regressão Logística
RN	Redes Neurais
ROC	<i>Receiver Operating Characteristic</i>
SVM	<i>Support Vector Machines</i>

Capítulo 1

2 Introdução

1.1 Contexto/Enquadramento

4 O retalho tem como objetivo a aquisição de bens ou serviços com o intuito de os revender
ao consumidor. Estas empresas não têm como função produzir ou transformar os bens
6 adquiridos, mas sim revendê-los ao consumidor. Podem assumir diversos formatos e dimensões
(hipermercado, loja de conveniência, supermercado, etc), localização (urbana, rural, etc),
8 estrutura legal (*franchising*, independente, etc), e comercializar diversos tipos de produtos
(alimentos, roupa, fármacos, etc). Podem ainda ser retalhistas especializados, comercializando
10 produtos relacionados com uma área específica, ou comercializar diversos tipos de produtos.

12 O retalho desempenha uma função vital na sociedade contemporânea. Com o crescente
aumento dos espaços urbanos, e com o aumento da população das cidades, cada vez menos
14 existem pessoas que tenham meios de subsistir sem adquirirem bens que são essenciais à vida.
É neste contexto que o retalho desempenha um papel importante ao providenciar produtos como
16 alimentos, roupa, entre outros, e também outros tipos de bens, que apesar de não serem essenciais
à vida, sem eles a vida quotidiana seria mais complicada. Já há alguns anos que as pessoas
18 dependem dos retalhistas para lhes providenciarem os bens que necessitam. O que tem mudado
ao longo dos últimos anos é a oferta disponível. Se antigamente o normal era a existência de
20 pequenos retalhistas que forneciam pequenas povoações, havendo pouca concorrência, hoje em
dia o cenário é bastante diferente. A área do retalho é mais competitiva do que em qualquer
22 outra época e é necessário ser melhor que a concorrência, de forma a sobreviver neste ambiente
tão competitivo. Por forma a obter algum destaque face à concorrência, muitas empresas de
24 retalho têm encontrado estratégias para cativar os clientes. Promoções, programas de
fidelização, publicidade são exemplos de ações com vista a atração de clientes. Este tipo de
26 iniciativas tem produzido excelentes resultados, mas o paradigma tem-se alterado um pouco.

Introdução

Antigamente todos os clientes eram tratados de igual forma pelas empresas, mas aos poucos as empresas aperceberam-se que existia a necessidade de conhecerem melhor os seus clientes, por forma a providenciar um serviço que vá ao encontro das necessidades dos seus clientes. Para tal surgiu a necessidade de recolher informação sobre os clientes. Atualmente as empresas recolhem informações dos mais variados tipos: idade, local de residência, nome e outros tipos de informação pessoal. A partir da informação recolhida, os retalhistas visam estudar o comportamento do cliente, compreender o padrão de ação e assim adaptar o serviço prestado por forma a satisfazer os diferentes tipos de clientes. Para além de informações sócio-demográficas, as empresas precisam de recolher informações sobre as suas interações com os clientes, nomeadamente sobre as transações realizadas.

A recolha de uma quantidade tão grande de informação tornou-se possível devido à evolução das tecnologias de informação. Estas também permitem analisar os dados e retirar informação importante, como padrões de comportamento. É possível compreender que tipo de produtos um determinado cliente adquire. É também possível verificar como é que os clientes reagem a campanhas promocionais que desenvolvem e com isto recolher informação fundamental para prever o comportamento dos clientes em futuras iniciativas do género. Assim as empresas conseguem fazer uma gestão mais eficiente dos seus recursos.

Esta tese será desenvolvida no contexto anteriormente referido, mais especificamente no contexto de uma empresa de retalho alimentar.

1.2 Motivação e Objetivos

A redução de custos é algo que qualquer empresa procura atualmente. Uma das origens de despesa para as empresas são as campanhas promocionais. Inúmeras vezes as campanhas realizadas não produzem os resultados esperados e investem-se quantidades significativas de recursos que acabam por não trazer um impacto significativo. Uma das motivações para a realização deste trabalho é prever a adesão dos clientes às campanhas promocionais e conseguir averiguar, antes de serem feitos grandes investimentos, se estas terão impacto significativo. Ao ser possível realizar uma previsão da adesão dos clientes é possível dirigir as campanhas aos clientes sensíveis às mesmas.

Este trabalho tem como objetivo o desenvolvimento de um modelo de previsão que permita averiguar a adesão dos clientes a uma determinada campanha. Para desenvolver este modelo serão utilizados algoritmos de aprendizagem. Para otimizar os resultados obtidos serão utilizados métodos *Ensemble*.

1.3 Contribuições

- 2 As contribuições desta tese são a aplicação de métodos *ensemble* na otimização de
modelos de previsão no contexto dos *cross-market discounts* e comparar a performance de
4 diferentes métodos *ensemble*. Outra contribuição é combinar os vários métodos *ensemble*
utilizados por forma a obter um modelo mais preciso.

6 1.4 Estrutura da Dissertação

- 8 Esta tese é constituída por 4 capítulos. No capítulo 2 será feito um enquadramento teórico
por forma a apresentar os conceitos abordados nesta tese. O capítulo 3 contém o caso de estudo,
onde é caracterizada a empresa em estudo, onde se faz uma análise dos dados e por fim é
10 descrita a metodologia utilizada para resolver o problema em mãos. Por fim, o capítulo 4
contém as principais conclusões tiradas.

Capítulo 2

2 Enquadramento Teórico

Neste capítulo serão analisados os conceitos relevantes no contexto desta tese. Na secção 2.1 será abordado o conceito de *Customer Relationship Management*, na secção 2.2 o conceito de *Data Mining*, na secção 2.3 o conceito de classificação e por fim na secção 2.4 o conceito de métodos *Ensemble*.

2.1 *Customer Relationship Management*

Durante muitos anos as empresas descoravam conhecer os seus clientes, tratando cada individuo de igual forma. Após a grande crise no sector da energia, surgiu a necessidade de aumentar a retenção de clientes promovendo uma mudança no paradigma empresarial. O foco das empresas passa a estar na retenção de clientes fazendo com que a relação entre o cliente e a empresa seja contínua, em vez da transação pontual (Sheth, 2002). Neste contexto, o *Customer Relationship Management* (CRM) ganhou proeminência.

Os objetivos do CRM passam por analisar o comportamento dos diversos clientes e adaptar as práticas empresariais por forma a maximizar a interação e o valor gerado por cada um destes. Para que este objetivo seja alcançado é crucial compreender cada cliente, individualmente (Grewal, Levy, and Kumar 2009). Isto torna-se especialmente relevante quando o valor das empresas está cada vez mais assente nos ativos intangíveis, tais como o conhecimento, marca, trabalhadores e também clientes (Gupta, Lehmann and Stuart, 2004). Boas práticas de CRM são exigidas às empresas por forma a maximizar o valor de cada cliente. O CRM possui quatro dimensões essenciais na obtenção de clientes (Ngai et al., 2009):

1. Identificação do cliente: nesta fase procura-se identificar os indivíduos que terão mais propensão para se tornarem clientes ou que trarão mais lucro para a empresa

Enquadramento Teórico

(Kracklauer et al., 2004). Para atingir este objetivo, as empresas têm adoptado práticas como a segmentação de clientes que consiste em subdividir os clientes em grupos com características semelhantes, e a identificação dos grupos que serão mais lucrativos para a empresa (Woo, Bae, & Park, 2005).

2. Atrair clientes: após identificar o público-alvo, o próximo passo é atrair o cliente. Isto é conseguido através de marketing direto, entrega cupões, correio, publicidade, etc.
3. Retenção do cliente: esta é a fase onde as práticas de CRM mais têm impacto. Kracklauer et al. (2004) indica que a satisfação do cliente é essencial. Esta aumenta quando a qualidade do serviço prestado vai de encontro ou supera as expectativas do cliente. Existem algumas práticas que aumentam a retenção dos clientes, como o marketing individual, programas de lealdade e gestão de reclamações.
4. Desenvolvimento do cliente: esta fase consiste em aumentar o valor do cliente para a empresa, isto é, o número de transações realizadas, o valor das transações e o lucro gerado.

No início dos anos 2000 estava em voga o princípio de que empresas em crescimento deveriam investir fortemente na aquisição de novos clientes (The Wall Street Journal, Nov 22,1999). Demers and Lev (2001) validaram este princípio em empresas que comercializavam produtos ou serviços através da Internet. Concluíram que fatores como o número de clientes distintos que visitavam o *site* e a sua capacidade de reter clientes estavam diretamente associados ao valor das ações das empresas. Mais recentemente, Gupta et.al (2004) concluíram que a capacidade de retenção aumenta o valor dos clientes para a empresa, sendo que um aumento de 1% na retenção poderá significar um aumento entre 3 a 7% no valor dos clientes. Um dos principais fatores que influencia a retenção é a satisfação do cliente (Gustafsson et al., 2005; Kracklauer et al.,2004). Esta é definida como a avaliação que o cliente faz do serviço prestado pela empresa, até à data (Johnson and Fornell 1991). Uma experiência positiva é alcançada quando a diferença entre as expectativas do cliente e a qualidade do serviço é reduzida. Assim sendo, a estratégia das empresas pode passar por reduzir as expectativas do cliente ou então por melhorar a qualidade do serviço (Boulding, 1993). Ao aumentar a satisfação do cliente as empresas beneficiam de várias vantagens. Rust e Zahorik(1993) concluíram que a satisfação dos clientes tem impacto na lealdade, na cota de mercado e nos lucros das empresas. Algumas empresas têm procurado promover a fidelização de clientes com recurso ao cartão de fidelização. Bolton et al.(2000) afirmam que clientes que aderem a programas de fidelização têm menos tendência a mudar para a concorrência quando esta possui uma oferta melhor. Além disso, a utilização do cartão de cliente permite às empresas recolher dados que lhes permitirão personalizar a sua abordagem, para cada cliente. Posteriormente, as empresas poderão recorrer a tecnologias de informação por forma a obter os padrões de comportamento dos diferentes clientes.

Vários são os meios aos quais as empresas têm recorrido para atrair, reter e desenvolver a relação com os clientes. Uma das formas é a realização de campanhas promocionais (Neslin, 2002). Os *Cross Market Discounts* são um tipo de promoções que têm estado em voga recentemente. Este tipo de promoção consiste em oferecer descontos em produtos ou serviços comercializado por uma empresa que atue num mercado não concorrente. Este tipo de descontos é obtido quando um cliente compra um certo tipo de produtos ou atinge um certo valor de compras efetuadas. Tem-se verificado que a implementação deste tipo de promoções permite aumentar as receitas. Clientes que estão próximos de cumprir os requisitos necessários para obter a promoção têm mais probabilidade de fazer uma nova compra (Sen, 2011). O lucro obtido através deste tipo de promoção aumenta com a importância de consumo, isto é, a necessidade que os clientes têm de obter os produtos ou serviços prestados pelas empresas. Mercados que apresentam grande frequência de consumo tem uma melhor chance de beneficiar com este tipo de descontos. O retalho alimentar e gasolineiras são dois negócios que beneficiam grandemente com este tipo de promoção visto que não competem diretamente e oferecem produtos com grande necessidade de consumo (Goic et al., 2011).

2.2 Extração de Conhecimento

O avanço nas tecnologias de informação tem permitido às empresas armazenar grandes quantidades de dados. Extraír daí conhecimento é um dos problemas que as empresas enfrentam. O armazenamento de grandes quantidades de dados, de clientes por exemplo, terá tanto valor quanto a qualidade da informação que for possível extrair e aferir. Surge então a necessidade de aplicar técnicas de extração de conhecimento. A princípio esta tarefa foi difícil de realizar tendo em conta que os algoritmos tinham sido desenvolvidos em laboratório, onde a quantidade de dados era reduzida e a qualidade garantida. Para resolver este problema surgiu a necessidade de criar um processo sistemático de preparação dos dados, e consequentemente aumentar a confiança nos resultados produzidos. Este processo é intitulado de extração de conhecimento em base de dados, e foi discutido pela primeira vez em 1989, num *workshop* do tema (Oliveira, 2012).

Na figura 1 estão ilustradas as etapas do processo de extração de conhecimento. A armazenagem dos dados é o primeiro passo do processo de extração de conhecimento. Após estar definido o objetivo pretendido, é necessário fazer uma seleção dos dados. O processo poderá passar pela seleção de uma amostragem dos dados ou conjunto de variáveis relevantes. O próximo passo será a limpeza dos dados. O processo começa com a limpeza de ruído ou de *outliers*, escolha de estratégias para lidar com dados em falta, e por fim, a definição da estrutura

mais conveniente para o posterior trabalho a desenvolver. A redução dos dados é a próxima tarefa a executar. A sua finalidade é agregar os dados em novas variáveis, aplicando métodos de transformação, criando assim variáveis que sejam mais significativas para o problema. Seguidamente é necessário escolher o tipo de tarefa de *data mining* a realizar (por exemplo classificação, regressão, etc). O passo que se segue é a escolha dos algoritmos a aplicar. Esta deve ser feita tendo em conta o objetivo do trabalho. Árvores de decisão, redes neuronais e regressão logística são alguns exemplos de algoritmos de classificação. Aplicando os algoritmos, é possível *à posteriori*, encontrar padrões que possam ser de interesse para o agente de decisão. Os últimos dois passos são a interpretação dos modelos obtidos e a aplicação do conhecimento extraído (Fayyad et al., 1996).

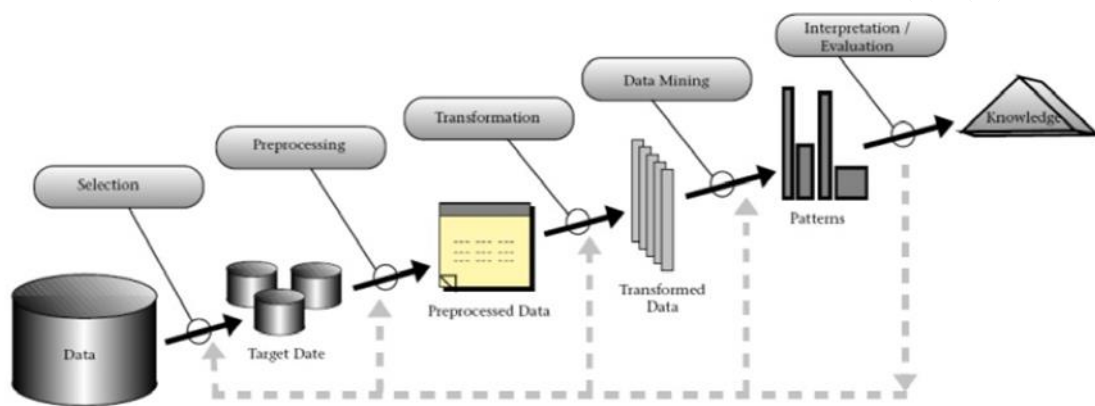


Figura 1 - Etapas do processo de extração de conhecimento de base de dados (Fayyad et al., 1996)

2.3 Data Mining

Data Mining consiste no processo de explorar e analisar grandes conjuntos de dados com o objetivo de encontrar padrões e regras que poderão ser importantes para apoiar a tomada de decisão (Berry and Linoff, 2000).

Existem essencialmente sete objetivos associados à utilização de técnicas de *Data Mining*:

1. Associação: visa estabelecer relações entre itens que coexistem num determinado registo (Mitra et al., 2002)
2. Classificação: classifica um elemento e insere-o numa de várias categorias predefinidas (Mitra et al., 2002).
3. *Clustering*: tem como objetivo criar grupos de clientes com base numa amostra heterogénea. Estes grupos são denominados de *clusters*, e ao contrário da classificação,

não são definidos *à priori*. Estes *clusters* são constituídos por registos com características semelhantes (Mitra et al., 2002)

4. *Forecasting*: visa estimar o valor futuro com base nos padrões dos registos. É aplicado quando os valores dos resultados são contínuos (Ahmed, 2004; Berry & Linoff, 2004).
5. Regressão: consiste em analisar os dados por forma a determinar o valor de um determinado registo (Carrier & Povel, 2003; Mitra et al., 2002).
6. *Sequence Discovery*: consiste em identificar padrões ou associações ao longo do tempo (Carrier & Povel, 2003; Mitra et al., 2002).
7. Visualização: consiste em apresentar os dados ao utilizador, de forma gráfica, com o intuito de identificar padrões. Alguns exemplos dos modelos de visualização são gráficos em 3D, “Hygraphs” e “SeeNet” (Shaw et al., 2001)

Por vezes estas técnicas produzem modelos que são demasiado dependentes dos dados. Este problema é denominado de *overfitting*. Uma forma de contornar este problema é utilizar *k-fold cross-validation*. Este método consiste em dividir o conjunto de dados em *k* partes, mutualmente exclusivas e de forma aleatória, e utilizar *k-1* partes para treino e uma parte para teste. Repete-se este processo *k* vezes para que todas as partes tenham sido utilizadas para treino e para teste (Kohavi, 1995). A utilização de *cross-validation* permite reduzir o *overfit* dos modelos treinados.

2.4 Classificação

Nesta secção será abordado o conceito de classificação e as técnicas que são utilizadas ao longo desta tese, nomeadamente regressão logística, árvores de decisão, *Support Vector Machines* e redes neurais.

O objetivo da classificação é prever a classe de um determinado registo. Estas classes estão predefinidas (Weiss and Kulikowski 1991; Hand 1981). Por isso mesmo, a classificação é uma técnica de apredizagem supervisionada. Numa primeira fase é utilizado um conjunto de dados, com várias variáveis e com a classe a que corresponde cada registo. Com estes dados será construído um modelo de classificação (fase de treino). Após este passo, é utilizado este modelo para prever a que classe pertence cada registo de um conjunto de dados que se pretenda classificar.

Posteriormente serão descritas algumas das técnicas mais populares utilizadas na classificação: regressão logística, árvores de decisão, *Support Vector Machines* e redes neurais.

2.4.1 Regressão Logística

A regressão logística tem como objetivo determinar a probabilidade de um evento com base nas variáveis independentes, que poderão ser categóricas ou numéricas. Esta técnica prevê a probabilidade de acontecimento acontecer ou não. Assim sendo, dado um conjunto de variáveis o modelo prevê se o acontecimento ocorrerá ou não e também indica a probabilidade de isso acontecer. A regressão logística é utilizada quando a variável dependente é binária. A regressão logística toma a forma evidenciada na Equação 1.

$$\text{Log}\left(\frac{p}{p-1}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$

Equação 1 - cálculo da probabilidade de um acontecimento (Ayer et al., 2010)

X_1, X_2, \dots, X_n representam as n variáveis independentes para previsão, β_0 é uma constante e $\beta_1, \beta_2, \dots, \beta_n$ são os coeficientes de regressão associados a cada uma das variáveis. Estes coeficientes são calculados com base nos dados utilizados para treino do modelo. O coeficiente é indicador do peso que a variável tem para a previsão (Ayer et al., 2010).

Os principais motivos para o uso desta técnica são a fácil implementação e a obtenção rápida e robusta de resultados (Buckinx and Van den Poel, 2005).

2.4.2 Árvores de Decisão

As árvores de decisão representam uma técnica de classificação bastante simples de implementar. São denominadas por árvores de decisão devido à sua estrutura: são constituídas por uma raiz, nós e folhas. A raiz é o ponto de partida. Cada nó representa uma variável do conjunto de dados. Cada nó pode-se sub-dividir noutros nós ou folhas. As folhas assumem o valor da previsão, isto é, o valor que se pretende prever.

No final dos anos 70 e início dos anos 80, J. Ross Quilan desenvolveu o algoritmo de árvore de decisão ID3 (Iterative Dichotomiser). Mais tarde introduziu um outro algoritmo, o C4.5. Em 1984, um grupo de cientistas (L. Breiman, J. Friedman, R. Olshen, and C. Stone) introduziram um novo algoritmo, o CART (Han e Kamber, 2006). Estes algoritmos adoptam uma abordagem *greedy*, nas quais as árvores são construídas do topo para baixo, isto é, desde a raiz, que é o topo da árvore, até as folhas. Uma abordagem *greedy* consiste em escolher a melhor alternativa, em cada iteração, com base na heurística utilizada pelo algoritmo. A figura 2 exemplifica uma árvore de decisão. Esta árvore permite prever se um cliente vai adquirir ou não um computador. Os nós são os retângulos e representam as variáveis de classificação. Neste caso as variáveis utilizadas são a idade, se é ou não estudante e a avaliação do crédito. Com

base nestas variáveis é construída a árvore. Cada nó é subdividido nos vários valores que a variável pode assumir. Aquando desta divisão, se alguma destas ramificações permitir determinar a classificação do registo então essa ramificação dá origem a uma folha. Isto também se verifica quando não existem mais variáveis para expandir a árvore. Caso não seja possível classificar o registo, ramifica-se novamente o nó, escolhendo uma das variáveis que ainda não tenha sido utilizada e repete-se o processo.

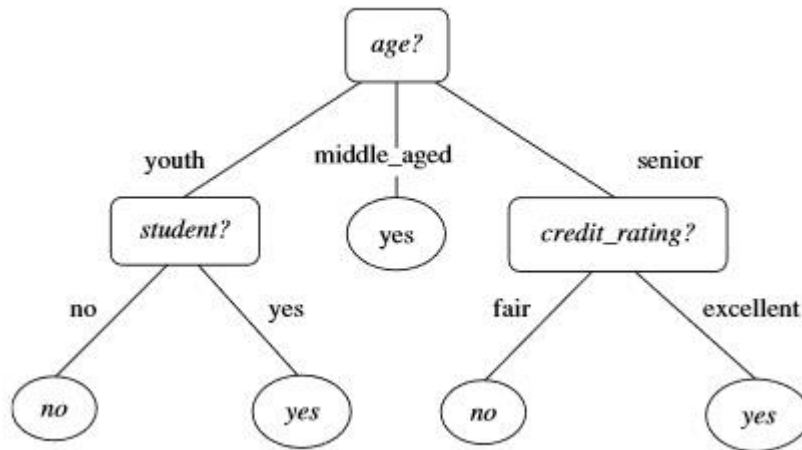


Figura 2 - Árvore de decisão (Han and Kamber, 2006)

2.4.3 Support Vector Machines

O algoritmo *Support Vector Machines* (SVM) consiste em mapear um conjunto de dados, com múltiplas dimensões, isto é, atributos, e encontrar um *hyperplane* que separe os pontos das diferentes classes. Este algoritmo é uma generalização não linear de um algoritmo desenvolvido na Rússia nos anos 60, denominado por *Generalized Portrait* (Smola e Schölkopf, 2004).

A figura 3 permite compreender este conceito. As cruzeiros e os círculos representam registos de classes distintas. Após se encontrar os pontos mais próximos de cada classe, que neste caso são os pontos dentro dos quadrados, procura-se o plano que maximiza a distância que divide as duas classes, com o menor número de objetos mal classificados. (Cortes e Vapnik, 1995).

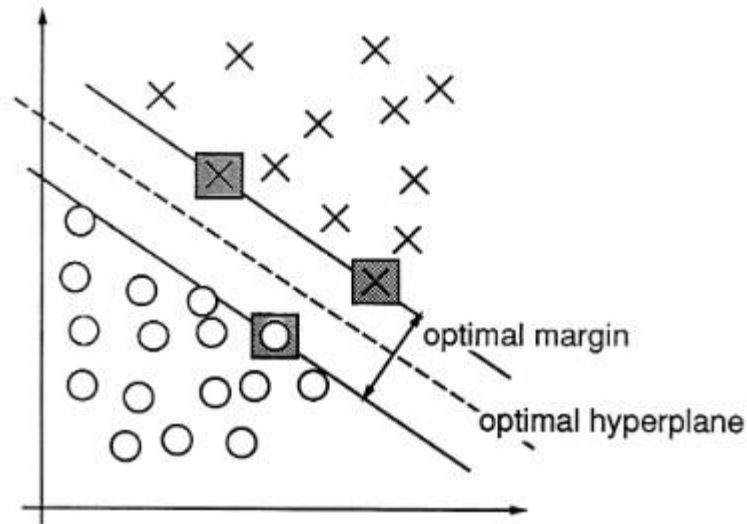


Figura 3 - Exemplo de um problema de duas dimensões
(Cortes e Vapnik, 1995)

2.4.4 Redes Neurais

Uma rede neuronal é constituída por camadas e neurónios, interligados entre si. A figura 4 ilustra o funcionamento de um neurónio. Cada entrada X é multiplicada por um peso W , podendo este ser positivo ou negativo. Existe um outro output denominado por *bias*, que é útil para regular a sensibilidade de um neurónio. O somatório dos vários *inputs* alimenta uma função de ativação que produz um valor de saída. As funções de ativação mais utilizadas são a *step*, *sign*, *sigmoid* e *linear*, funções essas que estão representadas na figura 5 (Negnevitsky, 2004).

Os neurónios estão interligados numa rede designada por topologia. Existem diversos tipos de topologias, sendo que estão inseridas em três categorias:

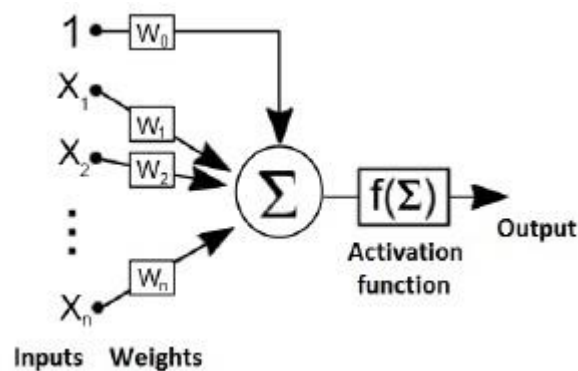


Figura 4 - Arquitetura de um neurónio

Enquadramento Teórico

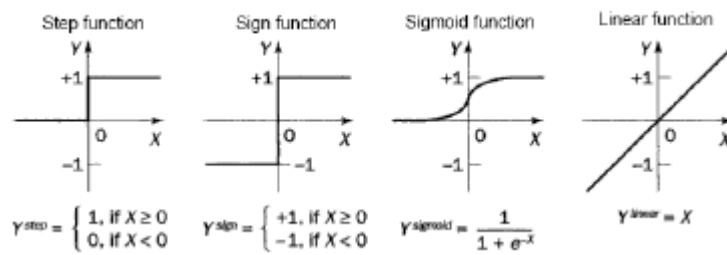


Figura 5 - Funções de ativação (Negnevitsky, 2004)

- 2 1. *Single layer feed-forward network*: neste tipo de topologia as entradas estão diretamente ligadas às saídas.

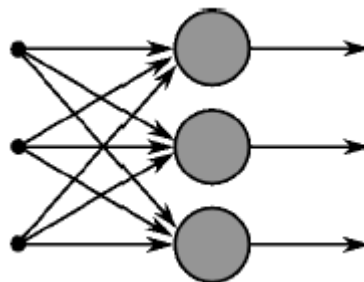


Figura 6 - Exemplo de *Single layer feed-forward network*

- 4 2. *Multiple layer feed-forward network*: esta topologia é constituída por várias
6 camadas. As camadas entre a camada de entrada e a camada de saída são
8 chamadas de *hidden layers*. Podem existir várias destas camadas, e o aumento do
número de *hidden layers* permite modelar funções mais complexas, mas ao mesmo
tempo tornando o processamento mais pesado.

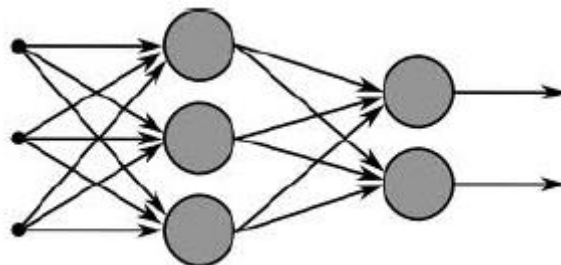


Figura 7 - Exemplo de *Multiple layer feed-forward network*

3. *Competitive or recurrent networks*: este tipo de topologia distingue-se das *feed-forward* pois apresenta pelo menos um ciclo. A inclusão destes ciclos faz com que a rede tenha um comportamento não linear.

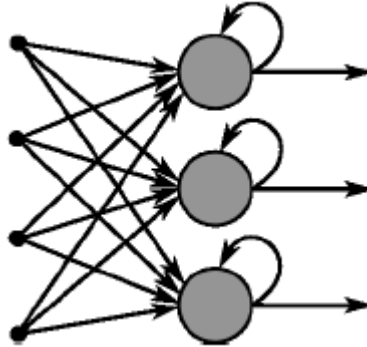


Figura 8 – Exemplo de *Competitive or recurrent networks*

2.5 Métodos *Ensemble*

Os métodos *ensemble* consistem em combinar um conjunto de classificadores individuais por forma a obter um melhor classificador. A investigação desenvolvida nesta área permite concluir que um *ensemble* é melhor que um classificador individual, na maioria dos casos. Vários têm sido os investigadores que se têm dedicado à investigação dos métodos *ensemble* (Breiman, 1996; Clemen, 1989; Perrone, 1993). Na investigação que tem sido realizada verificou-se que um bom *ensemble* é aquele em que os classificadores individuais são precisos e ao mesmo tempo erram em diferentes partes do conjunto de dados (Hansen & Salamon, 1990; Krogh & Vedelsby, 1995; Hashem, 1997; Opitz & Shavlik, 1996a, 1996b).

A figura 9 ilustra o conceito de ensemble. Neste exemplo representa-se um ensemble de redes neuronais, contudo é possível substituir as redes neuronais por outro classificador, como por exemplo, árvores de decisão. Cada uma das redes neuronais da figura (desde a O1 até à ON) será treinada com o conjunto de dados que lhe for atribuído. De seguida, a saída de cada rede (o1 até on) é combinada por forma a obter a saída do ensemble (\hat{o}). Vários investigadores defendem que uma forma eficaz de combinar os resultados dos classificadores é efetuar uma média das previsões (Alpaydin, 1993; Breiman, 1996c; Krogh & Vedelsby, 1995; Lincoln & Skrzypek, 1989).

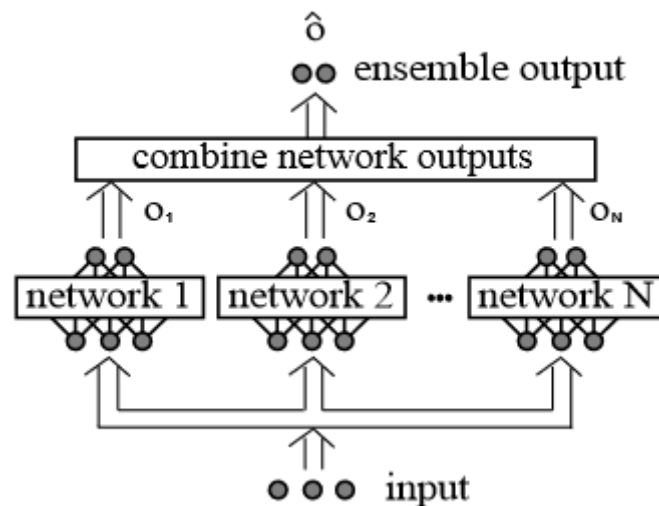


Figura 9 - Ensemble de redes neuronais (Maclin e Opitz, 1999)

Combinar o output de vários classificadores torna-se útil quando estes apresentem alguma discordância entre si. Combinar vários classificadores idênticos não apresenta qualquer tipo de benefício, pois o resultado obtido com o *ensemble* será muito semelhante ao dos classificadores individuais, não acrescentando assim grande valor.

Seguidamente serão analisados alguns tipos de *ensemble*. Os *ensembles* que serão analisados são os Random Forests, *bagging* e *boosting*.

2.5.1 Random Forests

A técnica *Random Forests* foi desenvolvida por Breiman (2001) e consiste num *ensemble* de árvores de decisão. Cada árvore de decisão é gerada utilizando diferentes partes do conjunto de dados. Para cada árvore é escolhido um conjunto de dados, aleatoriamente, e com reposição. Para além de utilizar diferentes dados, cada árvore de decisão utiliza um conjunto aleatório de atributos. O número de atributos seleccionados é, por norma, a raiz quadrada do número de atributos. Cada uma das árvores é expandida até ao limite, sem ser podada. Depois de serem construídas todas as árvores, é armazenada a classe que é atribuída a cada objeto, em cada árvore. Posteriormente é realizada uma votação para atribuir a cada objeto a sua classe (Oliveira, 2012).

As *Random Forests* apresentam várias vantagens. Como utilizam um vasto número de iterações, têm uma tendência muito reduzida para fazerem *overfit*. O facto de ser utilizada aleatoriedade torna esta técnica muito precisa em problemas de classificação e regressão (Breiman, 2001).

2.5.2 *Bagging*

O *bagging* foi desenvolvido por Breiman (1996). Para cada classificador é escolhido aleatoriamente e com reposição uma amostra dos dados. Esta amostra é, por norma, da mesma dimensão dos dados, mas como a amostragem é feita com reposição, vai haver registos que não estarão contemplados numa amostra e outros aparecerão duas ou mais vezes. São então gerados vários classificadores e o classificador final é a agregação de todos os classificadores gerados. Para classificar um registo é realizada uma votação, e é atribuída a classe com mais votos (Quinlan, 1996).

Por forma a obter melhorias nos resultados, é importante que os classificadores não sejam estáveis, e que pequenas mudanças no conjunto de dados tenham impacto nas previsões geradas, caso contrário os resultados do *bagging* poderão ser piores que os resultados dos classificadores individuais (Quinlan, 1996).

2.5.3 *Boosting*

Esta técnica consiste em combinar as previsões de vários classificadores do mesmo tipo e verificar em cada iteração o erro na previsão. Os registos que foram mal classificados terão um peso maior na próxima iteração. Assim, na iteração seguinte o algoritmo procura minimizar o erro de previsão, mas dará mais importância em prever corretamente os registos que apresentem um maior peso. Cada registo corretamente classificado vê o seu peso ser reduzido enquanto registos que foram incorretamente classificados vê o seu peso aumentar. Após várias iterações existem vários classificadores e é então realizada uma votação ponderada, na qual tem influência o erro na previsão do classificador (Freud e Schapire, 1996).

2.5.4 *Hybrid Ensemble*

O conceito de *hybrid ensemble* consiste em combinar vários *ensembles* por forma a obter um melhor classificador. Esta técnica geralmente apresenta melhorias relativamente aos *ensembles* individuais. Isto deve-se ao facto de diferentes algoritmos apresentarem melhor performance em diferentes casos, permitindo assim obter uma melhoria na performance global do algoritmo. Segundo Ballings (2014), o uso de seis sub-*ensembles* é suficiente para obter melhor performance do que o melhor sub-*ensemble* individual.

Existem duas maneiras de combinar as saídas dos vários classificadores: fusão e seleção. *Ensembles* que utilizam seleção procuram encontrar o classificador que melhor classifica uma instância em particular. Por outro lado, a fusão combina os resultados de todos os classificadores. Uma forma de fusão é a realização de uma votação, em que cada classificador vota na classe que determinou ser a correta (Gilpin e Dunlavy, 2008). Posteriormente, verifica-

se quantos votos tem a classe com mais votos e calcula-se a percentagem destes votos, sendo assim determinado o grau de certeza da previsão realizada.

2.6 Conclusão

Neste capítulo foram apresentados os conceitos base desta tese. Verificou-se que o paradigma empresarial mudou e que atualmente as empresas se preocupam mais em conhecer os seus clientes. As práticas de CRM têm evoluído por forma a atingir esse objetivo. São quatro as dimensões do CRM: identificação do cliente, atrair o cliente, retenção do cliente, e por fim, desenvolvimento do cliente. No que toca ao ponto de reter o cliente verificou-se que o aumento da satisfação do cliente desempenha um papel fundamental. Corresponder as expectativas do cliente levam a um aumento da satisfação. Analogamente, os programas de fidelização fazem com que o cliente tenha mais relutância em mudar para a concorrência, sendo necessária uma oferta muito apelativa para cativar a atenção do cliente e consequentemente leva-lo à mudança. Outra forma eficaz de atrair e reter clientes são a realização de promoções. Neste ponto verificou-se que os *cross-market discounts* são um tipo de promoção que está em voga. Consiste em oferecer descontos em produtos ou serviços comercializados por uma empresa que atue num mercado não concorrente.

Com os avanços tecnológicos realizados, é possível hoje em dia utilizar os computadores para aumentar o sucesso das campanhas promocionais das empresas, prevendo de que forma os clientes vão atuar perante uma determinada campanha. Para cumprir este objetivo são utilizadas técnicas de *data mining*. O *data mining* consiste no processo de explorar e analisar grandes conjuntos de dados com o objetivo de encontrar padrões e regras que poderão ser importantes para apoiar a tomada de decisão (Berry e Linoff, 2000). O processo de *data mining* passa por diversas fases: armazenamento de dados, seleção dos dados, limpeza dos dados, redução dos dados, escolha dos algoritmos, escolha do tipo de tarefa de *data mining*, aplicação dos algoritmos, interpretação dos modelos obtidos e por fim aplicação do conhecimento extraído. Existem essencialmente sete objetivos para as técnicas de *data mining*: associação, classificação, *clustering*, *forecasting*, regressão, *sequence discovery* e visualização.

Quando o problema que se tem em mãos é o de classificar registos, prevendo a sua classe a partir de um conjunto de classes predefinidas, está-se perante um problema de classificação. Neste tipo de problemas é utilizado um conjunto de dados, cuja classe é conhecida, para treinar um modelo que, posteriormente, permitirá prever a classe de registos não classificados. Algumas das técnicas mais populares utilizadas neste tipo de problemas são a regressão logística, árvores de decisão, *Support Vector Machines* e redes neuronais.

Enquadramento Teórico

2 Por forma a melhorar a performance dos algoritmos de *data mining* pode-se recorrer a
métodos *ensemble*. Os métodos *ensemble* consistem num conjunto de classificadores
individuais, que são combinados por forma a obter um melhor classificador. Os métodos
4 utilizados nesta tese são o *bagging*, *Random Forests* e *boosting*.

Capítulo 3

2 Caso de Estudo

3.1 Introdução

4 Prever a resposta dos clientes às campanhas promocionais tem sido uma das áreas de
grande investimento das empresas. A recolha de informação dos clientes tem tornado esta tarefa
6 possível. O estudo apresentado nesta tese surge no contexto do retalho alimentar e tem em vista
prever a resposta dos seus clientes a uma campanha promocional realizada em parceria com
8 uma empresa gasoleira.

3.2 Descrição da empresa

10 A empresa em estudo é um dos grandes retalhistas de alimentos a nível europeu. A sua
cadeia de lojas é constituída por hipermercados, lojas de conveniência e grandes supermercados.
12 Estas lojas diferenciam-se entre si na localização e na diversidade de produtos que oferecem.
Comercializam uma vasta gama de produtos, desde produtos alimentares, perecíveis, bazar
14 ligeiro e têxtil. No que toca aos produtos alimentares são comercializados produtos de mercearia
e bebidas. Na secção dos perecíveis comercializam produtos de talho, peixaria, padaria, frutas e
16 vegetais. Na área de bazar ligeiro a sua oferta vai desde produtos para a casa, livros, DVDs e
decoração. Na área têxtil comercialização roupa, calçado e acessórios.

18 A oferta variada da empresa e a sua forte presença no território em que opera têm-lhe
permitido manter-se na linha da frente na área do retalho alimentar. No entanto, a empresa
20 decide que existe a necessidade de se aproximar mais do cliente e conhecer um pouco melhor o
seu comportamento. Nesse sentido a empresa tem apostado em desenvolver o seu programa de
22 fidelização. Para tal a empresa criou o cartão de cliente. Este cartão permite ao cliente usufruir
de descontos, nas diversas lojas da empresa, e acumular o valor dos descontos em cartão para

utilizar numa futura compra. Com este cartão tornou-se possível recolher informação relativa ao cliente, como o nome, morada, data de nascimento, entre outros, e ainda recolher dados ao nível das transações como a data, o valor, produtos adquiridos e o local. Estes dados são armazenados em base de dados para posterior segmentação dos clientes. Estes dados são também utilizados, por exemplo, para apoiar o envio de descontos personalizados, com base nos padrões de compras dos clientes.

Para além das promoções oferecidas nos estabelecimentos comerciais da empresa, são também oferecidos descontos em combustível, que podem ser utilizados numa empresa gasolinera com a qual existe uma parceria. Para obter este desconto o cliente necessita de atingir um certo valor na transação e recebe um desconto que reduz o valor por litro de combustível. Este desconto é acumulado no cartão de cliente, que posteriormente poderá ser utilizado numa futura compra, na empresa de retalho.

3.3 Tratamento dos Dados

A primeira fase do processo de extração de conhecimento é o tratamento dos dados. Os dados utilizados são relativos às transações realizadas nas lojas da empresa, pelos seus clientes. Estes dados foram recolhidos entre Janeiro de 2011 e Junho de 2013. A campanha promocional com a gasolinera só entrou em vigor em Janeiro de 2013, sendo que os dados utilizados para o estudo do comportamento dos clientes são os dados recolhidos até ao final do ano de 2012. No total, a informação disponibilizada é constituída por mais de 64 milhões de registos, incluindo a seguinte informação:

1. Data da transação
2. ID do cliente
3. ID da transação
4. Loja
5. Referência do produto
6. Quantidade
7. Valor total (quantidade x preço individual)
8. Tipo de cliente (ocasional, frequente, leal)

Para ser possível retirar informação relevante sobre cada cliente foi necessário agregar a informação transacional. Numa primeira fase, os dados foram agregados ao nível da transação, por forma a determinar que clientes receberam os descontos em combustível. Numa segunda fase, os dados foram agregados ao nível do cliente. A partir dos dados referidos anteriormente foram geradas as seguintes variáveis:

1. ID Cliente – Identificação do cliente.

Caso de Estudo

2. Número de transações realizadas – Esta variável indica o número total de transações realizadas pelo cliente. Com esta variável pretende-se verificar se os clientes que aderem à campanha realizam mais transações que os não aderentes.
3. Total de artigos adquiridos – Esta variável indica o número total de artigos adquiridos. Com esta variável pretende-se compreender se os clientes aderentes compram mais artigos que os não aderentes.
4. Número médio de artigos adquiridos por transação – Esta variável indica o número de artigos que o cliente compra, em média, por transação. Com esta variável pretende-se verificar a relação entre o número médio de artigos comprados, por transação, e a propensão para aderir à campanha.
5. Número médio de artigos adquiridos por mês - Esta variável indica o número de artigos que o cliente compra, em média, por mês. Com esta variável pretende-se verificar se clientes que comprar um número médio de artigos, por mês, superior se têm mais propensão para aderir à campanha.
6. Tipo de cliente – Esta variável indica o tipo de cliente (ocasional, frequente, leal). Com esta variável pretende-se verificar qual o tipo de cliente que adere mais frequentemente à campanha.
7. Última transação realizada - Esta variável indica o número de dias passados desde a última transação efetuada. Com esta variável pretende-se verificar se os clientes que não efetuam transações há muito tempo aderem à campanha ou apenas os que efetuaram compras recentemente.
8. Primeira transação realizada - Esta variável indica o número de dias passados desde a primeira transação efetuada, no período de análise. Esta variável permite verificar se os clientes cuja atividade transacional durante o período de análise se iniciou mais cedo aderem mais frequentemente que os clientes cuja atividade se iniciou mais tarde.
9. Valor total dos artigos adquiridos – Esta variável indica o valor total de todas as transações. Com esta variável é possível verificar se os clientes que gastaram mais, ao longo do seu período de interação com a empresa retalhista, têm mais propensão para aderir à campanha, comparativamente com clientes que gastaram menos.
10. Valor médio gasto por transação – Esta variável indica o valor gasto, em média, por transação. Com esta variável é possível verificar se são os clientes que gastam mais dinheiro, em média, por transação, que são mais sensíveis à campanha promocional ou se são os clientes que dispendem menos.
11. Valor médio gasto por mês - Esta variável indica o valor gasto, em média, por mês. Com esta variável pretende-se verificar se o valor médio gasto por mês é um indicador da tendência de um cliente em aderir à campanha promocional.
12. Valor médio dos artigos adquiridos – Esta variável indica o valor médio dos produtos adquiridos, por um determinado cliente. Com esta variável é possível verificar se os

clientes aderentes comprem artigos de valor médio baixo ou se são clientes que adquirem produtos mais dispendiosos.

13. Utiliza talão? – Esta variável indica se o cliente aderiu ou não à campanha promocional.

Nos dados originais existem registos que indicam as transações realizadas pelo cliente na empresa gasoleira, onde foram utilizados os talões de desconto que tinham sido emitidos pela empresa de retalho. Com estes dados foi possível verificar quais os clientes que aderiram à campanha promocional. No total, a amostra de dados disponibilizada pela empresa contém 69.096 clientes, e destes 24% aderiram à campanha promocional.

Após a agregação dos dados foi realizada uma limpeza. Por forma a realizar esta limpeza foram utilizadas as variáveis de valor médio por transação e número médio de transações por mês, sendo estas as variáveis que apresentam os valores mais relevantes sobre o padrão de comportamento do cliente. Clientes cujos valores destas variáveis excedessem o intervalo definido pela média mais três desvios padrão foram excluídos, uma vez que se acredita não traduzirem o comportamento habitual de um consumidor final. Pretende-se desconsiderar da análise os pequenos retalhistas que se abastecem nas lojas da empresa em estudo. Após a limpeza, o número remanescente de cliente foi de 66.580, ou seja, cerca de 96% dos dados originais.

3.4 Análise Preliminar de Comportamentos

Os dados que serão analisados de seguida são referentes aos dados após a limpeza, ou seja, os dados de 66580 clientes.

Com base nos dados referidos anteriormente, foi possível verificar que em média, os clientes vão 3.71 vezes por mês às compras. Ao comparar a média de visitas à loja para os clientes aderentes verificou-se que os clientes que aderem à campanha vão às compras, em média, 5.26 vezes por mês, enquanto os clientes não aderentes vão apenas 3.23 vezes.

Outra variável de interesse é o valor médio. Em média, os clientes dispendem cerca de 29.8€, em cada transação. Quando é feita a média para aderentes e não aderentes, a média dos clientes aderentes é de 37.35€ enquanto a média dos clientes não aderentes é de 27.47€. Com base nestes dados, os clientes que aderem à campanha têm tendência a dispendem mais dinheiro, em média, do que os clientes que não aderem à campanha. Por fim, temos a variável de classificação dos clientes com base em segmentos de valor para a empresa, que se baseia na frequência de visitas à loja e valor das transações. A exploração destes segmentos está fora do

Caso de Estudo

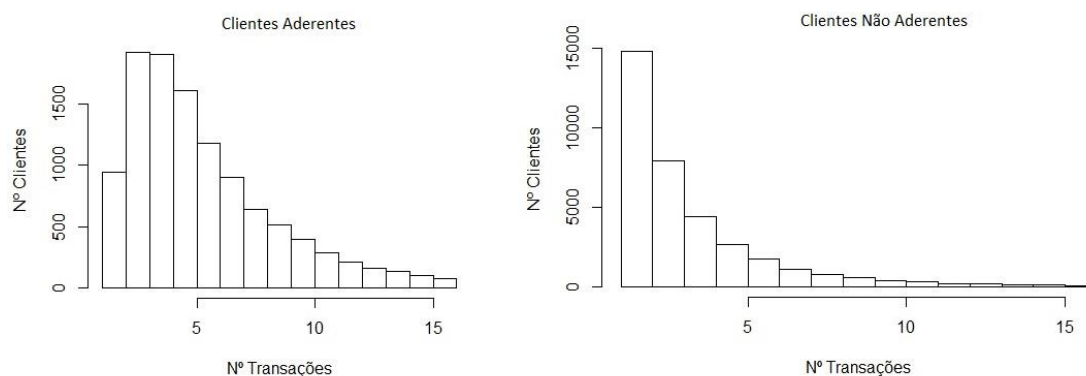


Figura 11 - Histogramas do Número Médio de Transações p/Mês

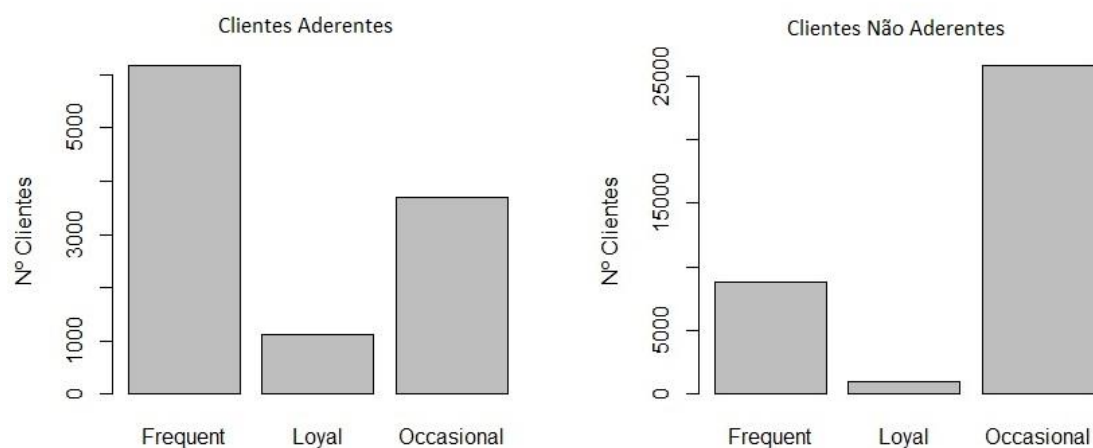


Figura 10 - Gráfico de barras relativo à classificação dos clientes

âmbito desta tese. No entanto, importa referir que a empresa de retalho classifica os seus clientes em três categorias: leal, frequente e ocasional. Com base na figura 11, é possível verificar que para os clientes aderentes (gráfico da esquerda) a percentagem de clientes frequentes (56%) e leais (10%) é superior à dos clientes não aderentes (frequentes: 25%; leais: 3%) e que a percentagem de clientes ocasionais é maior no lado dos clientes não aderentes, sendo estes 72%, face aos 34% do lado dos clientes aderentes (gráfico da direita). É possível concluir que os clientes classificados com as categorias “frequente” e “leal” têm mais propensão para aderir à campanha promocional.

Em resumo, os clientes que têm mais propensão para aderir à campanha promocional vão, em média, mais vezes às compras, dispendem mais dinheiro e são classificados como sendo “frequente” ou “leal”.

Caso de Estudo

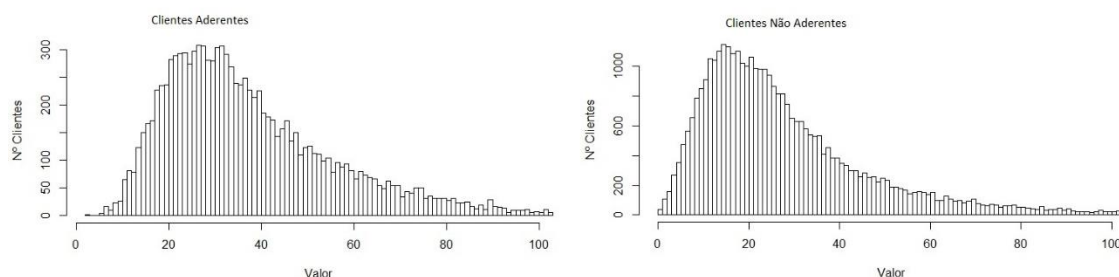


Figura 12 - Histogramas do Valor Médio p/Transação

Na tabela 1 encontram-se os valores médios para as restantes variáveis. Relativamente ao nº total de transações é possível verificar que os clientes aderentes realizam, em média, mais do dobro das transações que os clientes não aderentes. O número total de artigos também é maior no grupo dos clientes aderentes, sendo este valor 2,68 vezes maior que no grupo dos não aderentes. Tanto o número médio de artigos por transação, como por mês, são superiores no grupo dos aderentes. Com base na variável de nº de dias desde a última transação, verifica-se que, em média, o grupo dos clientes aderentes efetuou a sua última transação em dias mais próximos do dia de início da campanha promocional. O grupo dos clientes aderentes apresenta, em média, clientes cuja atividade transacional, durante o período de análise, se iniciou mais cedo. No que toca ao valor total gasto pelos clientes aderentes, este é bastante superior, em média, ao do grupo dos clientes não aderentes, sendo o valor do grupo dos aderentes 2,9 vezes superior. O valor médio gasto por mês também é superior no grupo dos clientes aderentes, sendo este 2,12 vezes superior. Por fim, o valor médio por artigo é ligeiramente superior no grupo dos não aderentes, sugerindo que estes clientes adquirem, em média, artigos de valor superior ao dos clientes aderentes.

Variável	Aderentes	Não Aderentes
Nº Total de Transações	111,89	52,92
Nº Total de Artigos	1947,30	726,13
Nº Médio de Artigos p/Transação	18,09	13,51
Nº Médio de Artigos p/Mês	88,55	42,77
Nº Dias desde a Última Transação	11,04	87,24
Nº Dias desde a Primeira Transação	668,53	589,03
Valor Total das Transações	5000,68	1722,44
Valor Médio p/Mês	179,99	84,88
Valor Médio dos Artigos	2,72	2,89

Tabela 1 - Valor médio das variáveis para clientes aderentes e não aderentes

3.5 Metodologia

Nesta sub-secção será abordada a metodologia utilizada para a construção dos modelos. Em primeiro lugar foram definidas as variáveis, com base nas transações dos clientes, sendo estas as variáveis de previsão, utilizadas para compreender o tipo de clientes que aderem à campanha. A variável *target* consiste numa variável binária, que indica se o cliente adere ou não à campanha promocional.

Para cada modelo foi utilizado *10-fold cross-validation*. Esta metodologia consiste em dividir os dados em 10 partes semelhantes e utilizar 9 partes para treino do modelo e uma parte para teste. Todas as partes são utilizadas como treino e teste. São gerados 10 modelos, em que cada um dos modelos é testado com uma parte diferente e treinado com as 9 partes restantes. Por fim é feita uma média das previsões de cada modelo por forma a obter o modelo final. As 10 partições foram geradas e posteriormente foi realizado o treino dos modelos, com as mesmas partições, por forma a ser possível comparar a performance de cada modelo. Em primeiro lugar foram contruídos os modelos utilizando as técnicas de *data mining*, individualmente. As técnicas utilizadas foram as árvores de decisão, redes neuronais, *Support Vector Machines* e regressão logística. De seguida foram contruídos modelos de previsão utilizando métodos *ensemble*. Os métodos *ensemble* utilizados foram o *bagging*, *adaboost* e *Random Forests*. Relativamente ao *bagging*, foram contruídos três modelos, utilizando técnicas de *data mining* diferentes. As três técnicas utilizadas foram as redes neuronais, *Support Vector Machines* e a regressão logística. Os métodos *ensemble adaboost* e *Random Forests* foram contruídos utilizando árvores de decisão.

Foi utilizado o *software* R como ferramenta auxiliar para construção dos modelos. Os parâmetros utilizados nas árvores de decisão foram a dimensão mínima do nó a dois e o parâmetro da dimensão mínima de cada filho a um (para mais informação ver: Han e Kamber, 2006). Relativamente às redes neuronais, o número de neurónios na *hidden layer* utilizado foi seis. Este número foi determinado por forma a que o número de *hidden neurons* esteja entre a dimensão da camada de entrada e a de saída (Heaton, 2008). Tendo em conta que a dimensão da camada de entrada é doze e a de saída é um, seis seria o valor mais recomendado a utilizar. O número máximo de iterações foi fixado em mil e o *weight decay* foi de 0.001 (para mais informação ver: Negnevitsky, 2004). Relativamente aos SVM foi utilizado o *kernel radial basis*, com o parâmetro gamma a assumir o valor 2^{-1} e o parâmetro C o valor de 2^4 (para mais informação ver: Cortes e Vapnik, 1995). A regressão logística foi implementada fixado-se o valor do número máximo de iterações em 50 (para mais informação ver: Ayer et al., 2010).

Na função de *bagging*, o número de iterações realizadas foi de 10, e em cada iteração foi utilizada uma amostra de dados com a dimensão do conjunto de dados original, sendo que esta

amostra foi escolhida de forma aleatória, com reposição. As amostras utilizadas foram as mesmas para as três implementações de *bagging*. Em cada uma destas implementações, os parâmetros utilizados pelos algoritmos de aprendizagem foram os mesmos que os parâmetros utilizados na construção dos modelos individuais. Na implementação do método *adaboost*, o número de árvores utilizado foi de 50. Os parâmetros das árvores de decisão utilizadas pelo algoritmo *adaboosts* são os mesmos utilizados na implementação das árvores de decisão, referida anteriormente. Por fim, as *Random Forests* foram treinadas com 500 árvores.

3.6 Resultados

Para avaliar a performance dos modelos, foi utilizado o método de avaliação *Area Under the Curve* (AUC). Este método determina a área da curva *Receiver Operating Characteristic* (ROC). O valor da área varia entre 0 e 1. Este valor indica a probabilidade de o modelo fazer uma previsão correta, sendo que valores próximos de 1 indicam que o modelo tem uma maior probabilidade de fazer uma previsão correta (Bradley, 1997).

Na tabela 1 estão os valores de AUC para cada técnica. Em primeiro lugar foram desenvolvidos os modelos utilizando as técnicas de *data mining* individualmente. A análise da tabela permite verificar que a técnica que apresenta melhores resultados, neste caso de estudo, é a regressão logística. As Árvores de Decisão apresentam uma performance ligeiramente pior, tal como os *Support Vector Machines*. As Redes Neurais apresentaram a pior performance. Em nenhuma destas técnicas foi realizado o *tuning* dos parâmetros, sendo que a performance individual obtida poderá ser superior com uma optimização destes. De seguida, foram implementados métodos *ensemble* por forma a optimizar a performance dos algoritmos de *data mining*. Os métodos *ensemble* implementados que utilizam árvores de decisão são o *adaboost* e as *Random Forests*. Com base na tabela 2 é possível verificar que o algoritmo *adaboost* melhorou a performance das árvores de decisão (AUC difere em 0.0341) e as *Random Forests* também melhoraram a performance (AUC difere em 0.048). A rede neuronal foi o algoritmo que apresentou a pior performance dos algoritmos de *data mining*. Após aplicar o algoritmo *bagging*, verificou-se uma melhoria na performance (AUC difere em 0.2335). Relativamente ao algoritmo *Support Vector Machines*, verificou-se uma melhoria, após a aplicação do algoritmo *bagging* (AUC difere em 0.0207). Por fim, a regressão logística apresentou uma melhoria marginal de performance, sendo este o algoritmo que menos beneficiou da aplicação de métodos *ensemble*. Por fim, foi aplicado o algoritmo *hybrid ensemble*, que neste caso consistiu em realizar uma votação com os vários modelos construídos, através de métodos *ensemble*, por forma a obter um modelo melhor. Neste caso, a melhoria de performance, relativamente ao método *ensemble* com melhor performance traduziu-se num aumento da AUC em 0.024. Assim sendo, é possível verificar que a aplicação de métodos *ensemble* permitiu melhorar a

performance relativamente à aplicação das técnicas individualmente. Ao combinar os *ensembles* com a técnica de *hybrid ensemble* foi possível obter uma melhoria relativamente a cada um dos métodos *ensemble* individuais.

Técnica	AUC
Árvores Decisão (AD)	0.8077
Redes Neurais (RN)	0.6132
<i>Support Vector Machines</i> (SVM)	0.7951
Regressão Logística (RL)	0.8469
<i>Bagging</i> RN	0.8467
<i>Bagging</i> SVM	0.8158
<i>Bagging</i> RL	0.8472
<i>Adaboost</i>	0.8418
<i>Random Forests</i>	0.8555
<i>Hybrid Ensemble</i>	0.8795

Tabela 2 - AUC de cada técnica implementada

3.7 Discussão

O tipo de campanhas em estudo, os *cross-market discounts*, requerem das empresas envolvidas grandes esforços para que se tornem viáveis. Isto deve-se ao facto de existir trocas de informação entre dois sistemas de informação diferentes, pensados para satisfazer as necessidades que cada uma das empresas tem. Como tal, nem sempre é simples a interligação de sistemas e é necessário investir várias horas de trabalho para conceber uma forma de comunicação e troca de informação. O tipo de modelo desenvolvido, no contexto desta tese, permite às empresas prever o sucesso de campanhas deste género, e assim ter um auxílio na tomada de decisão de implementar outras iniciativas semelhantes no futuro, que poderá fazer poupar às empresas uma grande quantidade de recursos. Este tipo de modelos poderá também servir para direccionar os descontos apenas aos clientes com propensão para usufruírem do desconto. Neste caso em particular, estes modelos poderão ser uma grande ajuda tendo em conta que apresentam performances elevadas, dando grande confiança nas previsões que fornecem. Assim sendo, a tomada de decisão, com base nas previsões obtidas, está bem fundamentada e com um elevado grau de certeza.

3.8 Conclusão

- 2 Neste capítulo foi abordado o processo de extração de conhecimento do caso de estudo.
Foi possível verificar que os clientes que aderem às campanhas promocionais são os clientes
4 que vão, em média, mais vezes às compras, dispendem mais dinheiro e são classificados como
sendo “frequente” ou “leal”.
- 6 Relativamente às técnicas de *data mining* verificou-se que os métodos *ensemble*
apresentam melhorias relativamente às técnicas de *data mining*, quando aplicadas
8 individualmente. Por fim combinar as previsões dos métodos *ensemble* com a técnica *hybrid*
ensemble permitiu obter uma melhor performance relativamente aos *ensemble* individuais.

Capítulo 4

2 Conclusões e Trabalho Futuro

4.1 Satisfação dos Objectivos

4 O objetivo principal desta tese é a construção de modelos que permitam à empresa de
6 retalho, usada como caso de estudo, prever a reação dos seus clientes à campanha promocional
8 que lançou em conjunto com uma empresa gasolinera. Esse objetivo foi cumprido, visto que os
10 modelos foram construídos e apresentam performances elevadas, sendo que para os métodos
12 *ensemble* todos eles apresentam uma probabilidade de realizar uma previsão corretamente
superior a 80%. Assim sendo, estes modelos poderão auxiliar a empresa de retalho a tomar
decisões sobre esta campanha, nomeadamente na definição do seu *target*, possibilitando uma
redução nos custos associados a esta ou futuras campanhas.

14 O trabalho também permite compreender que tipo de clientes que são mais propensos a
16 aderir às campanhas. Verificou-se que os clientes que têm mais tendência a aderir são clientes
que, em média, vão mais vezes por mês às compras, dispendem mais dinheiro e são
classificados pela empresa de retalho como sendo clientes frequentes ou leais. No grupo dos não
aderentes, mais de metade são clientes ocasionais.

18 Relativamente às técnicas de *machine learning*, verificou-se que as técnicas de *data*
20 *mining*, quando utilizadas individualmente, como por exemplo árvores de decisão, redes
neurais, e outras, apresentam uma probabilidade de efetuar uma previsão correta inferior aos
22 métodos *ensemble*. Qualquer uma das técnicas *ensemble* utilizadas, como o *bagging*, *adaboost* e
Random Forests, apresentaram melhorias significativas na performance dos modelos.

4.2 Limitações / Trabalho Futuro

As limitações do trabalho desenvolvido passam pelo acesso limitado aos dados dos clientes. Num trabalho futuro seria interessante considerar os dados sócio demográficos dos clientes e compreender de que forma estes têm impacto nos modelos de previsão. Outra melhoria será a utilização de um número maior de clientes.

Num trabalho futuro seria interessante explorar outro tipo de iniciativas, como outro tipo de promoções ou campanhas. Outro aspeto a trabalhar no futuro é a utilização de outras técnicas de *data mining* e também diferentes métodos *ensemble*.

Outra limitação é o facto de não terem sido removidos os clientes que não receberam cupões do estudo. Num trabalho futuro será de interesse retirar estes clientes e verificar apenas o comportamento dos clientes que receberam cupões. Outro aspecto interessante a explorar é a construção de um modelo que permita estimar o número de cupões utilizados e não a avaliação de utilização ou não.

Referências

Alpaydin, E. (1993). Multiple networks for function learning. In Proceedings of the 1993 IEEE International Conference on Neural Networks, Vol. I, pp. 27–32 San Francisco.

Ahmed, S. R. (2004). Applications of data mining in retail business. *Information Technology: Coding and Computing*, 2, 455–459.

Ayer, T., Chhatwal, J., Alagoz, O., Kahn Jr, C. E., Woods, R. W., & Burnside, E. S. (2010). Comparison of Logistic Regression and Artificial Neural Network Models in Breast Cancer Risk Estimation 1. *Radiographics*, 30(1), 13-22.

Ballings, Michel. Advances and applications in Ensemble Learning. Diss. Faculty of Economics and Business Administration Advances and Applications in Ensemble Learning, Michel Ballings 2014 advisor: Prof. DR. Dirk Van Den Poel Dissertation submitted to the Faculty of Economics and Business Administration, Ghent University, 2014.

Berry, M. J. A. and Linoff, G. S. (2000). *Mastering Data Mining: The Art and Science of Customer Relationship Management*. Wiley, New York, 1 edition.

Berry, M. J. A., & Linoff, G. S. (2004). *Data mining techniques second edition - for marketing, sales, and customer relationship management*. Wiley.

Bolton, Ruth N., P. KandBramlettMatthewD Kannan, and Matthew D. Bramlett. "Implications of loyalty program membership and service experiences for customer retention and value." *Journal of the academy of marketing science* 28.1 (2000): 95-108.

Bose, Ranjit. "Customer relationship management: key components for IT success." *industrial management & data systems* 102.2 (2002): 89-97.

Boulding, William, et al. "A dynamic process model of service quality: From expectation to behavioral intentions." *Journal of Marketing Research*, Vol 30 (1993): 7-27.

Bradley, Andrew P. "The use of the area under the ROC curve in the evaluation of machine learning algorithms." *Pattern recognition* 30.7 (1997): 1145-1159.

Referências

- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123–140.
- Breiman, L. (1996). Stacked regressions. *Machine Learning*, 24(1), 49–64.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1):5–32.
- Buckinx, W. and Van den Poel, D. (2005). Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting. *European Journal of Operational Research*, 164(1):252–268
- Clemen, R. (1989). Combining forecasts: A review and annotated bibliography. *Journal of Forecasting*, 5, 559–583
- Cortes, Corinna, and Vladimir Vapnik. "Support-vector networks." *Machine learning* 20.3 (1995): 273-297.
- Demers, Elizabeth and Baruch Lev (2001), "A Rude Awakening: Internet Shakeout in 2000," *Review of Accounting Studies*, 6, 2/3, 331-359.
- Fayyad, Usama, Gregory Piatetsky-Shapiro, and Padhraic Smyth. "The KDD process for extracting useful knowledge from volumes of data." *Communications of the ACM* 39.11 (1996): 27-34.
- Freund, Yoav, and Robert E. Schapire. "Experiments with a new boosting algorithm." *ICML*. Vol. 96. 1996.
- GILPIN, SEAN A., and DANIEL M. DUNLAVY. "Heterogeneous Ensemble Classification." *CSRI SUMMER PROCEEDINGS 2008* (2008): 90.
- Giraud-Carrier, C., and O. Povel. "Characterising data mining software." *Intelligent Data Analysis* 5 (2001): 1-12.
- Goic, Marcel, Kinshuk Jerath, and Kannan Srinivasan. "Cross-market discounts." *Marketing Science* 30.1 (2011): 134-148.
- Gupta, Sunil, Donald R. Lehmann, and Jennifer Ames Stuart. "Valuing customers." *Journal of marketing research* 41.1 (2004): 7-18.

Referências

Gustafsson, Anders, Michael D. Johnson, and Inger Roos. "The effects of customer satisfaction, relationship commitment dimensions, and triggers on customer retention." *Journal of marketing* 69.4 (2005): 210-218.

Han, J. K., & Kamber, M. M.(2006) *Data mining: concepts and techniques*.

Hand, D. J. 1981. *Discrimination and Classification*. Chichester, U.K.: Wiley.

Hansen, L., & Salamon, P. (1990). Neural network ensembles. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12, 993–1001.

Hashem, S. (1997). Optimal linear combinations of neural networks. *Neural Networks*, 10(4), 599–614.

Heaton, Jeff. *Introduction to neural networks with Java*. Heaton Research, Inc., 2008.

Hsu, Chih-Wei, Chih-Chung Chang, and Chih-Jen Lin. "A practical guide to support vector classification." (2003).

Johnson, Michael D. and Claes Fornell (1991), "A Framework for Comparing Customer Satisfaction Across Individuals and Product Categories," *Journal of Economic Psychology*, 12 (2), 267–86.

Kohavi, Ron. "A study of cross-validation and bootstrap for accuracy estimation and model selection." *Ijcai*. Vol. 14. No. 2. 1995.

Kracklauer, A. H., Mills, D. Q., and Seifert, D. (2004). Customer management as the origin of collaborative customer relationship management. In *Collaborative customer relationship management: taking CRM to the next level*. Springer.

Krogh, A., & Vedelsby, J. (1995). Neural network ensembles, cross validation, and active learning. In Tesauro, G., Touretzky, D., & Leen, T. (Eds.), *Advances in Neural Information Processing Systems*, Vol. 7, pp. 231–238 Cambridge, MA. MIT Press.

Lincoln, W., & Skrzypek, J. (1989). Synergy of clustering multiple back propagation networks. In Touretzky, D. (Ed.), *Advances in Neural Information Processing Systems*, Vol. 2, pp. 650–659 San Mateo, CA. Morgan Kaufmann.

Referências

Maclin, Richard, and David Opitz. "Popular ensemble methods: An empirical study." *Journal of Artificial Intelligence Research* (1999).

Mitra, Sushmita, Sankar K. Pal, and Pabitra Mitra. "Data mining in soft computing framework: a survey." *IEEE transactions on neural networks* 13.1 (2002): 3-14.

Negnevitsky, M. (2004). *Artificial Intelligence: A Guide to Intelligent Systems*. Addison Wesley, Boston, 2 edition.

Neslin, S. A. 2002. *Sales Promotion*. Marketing Science Institute, Cambridge, MA.

Ngai, Eric WT, Li Xiu, and Dorothy CK Chau. "Application of data mining techniques in customer relationship management: A literature review and classification." *Expert systems with applications* 36.2 (2009): 2592-2602.

Oliveira, Vera Lúcia Miguéis. *Analytical Customer Relationship Management in Retailing Supported by Data Mining Techniques*. Diss. Universidade do Porto, 2012.

Opitz, D., & Shavlik, J. (1996a). Actively searching for an effective neural-network ensemble. *Connection Science*, 8(3/4), 337–353.

Opitz, D., & Shavlik, J. (1996b). Generating accurate and diverse members of a neural-network ensemble. In Touretsky, D., Mozer, M., & Hasselmo, M. (Eds.), *Advances in Neural Information Processing Systems*, Vol. 8, pp. 535–541 Cambridge, MA. MIT Press.

Perrone, M. (1993). *Improving Regression Estimation: Averaging Methods for Variance Reduction with Extension to General Convex Measure Optimization*. Ph.D. thesis, Brown University, Providence, RI.

Prinzie, Anita, and Dirk Van den Poel. "Constrained optimization of data-mining problems to improve model performance: A direct-marketing application." *Expert Systems with Applications* 29.3 (2005): 630-640.

Quinlan, J. Ross. "Bagging, boosting, and C4. 5." *AAAI/IAAI*, Vol. 1. 1996.

Rust, Roland T., and Anthony J. Zahorik. "Customer satisfaction, customer retention, and market share." *Journal of retailing* 69.2 (1993): 193-215.

Sen, Boudhayan. "Customer Response To Cross-Market Rewards." (2011).

Referências

Shaw, M. J., Subramaniam, C., Tan, G. W., & Welge, M. E. (2001). Knowledge management and data mining for marketing. *Decision Support Systems*, 31, 127–137.

Sheth, Jagdish N. "The future of relationship marketing." *Journal of services marketing* 16.7 (2002): 590-592.

Smola, Alex J., and Bernhard Schölkopf. "A tutorial on support vector regression." *Statistics and computing* 14.3 (2004): 199-222.

Tan, P., Steinbach, M., and Kumar, V. (2006). *Introduction to Data Mining*. Addison Wesley, Boston, 1 edition.

The Wall Street Journal, "Buying the Buyers: The goal these days seems to be to attract customers, whatever they cost you," Nov 22, 1999, B1.

Weiss, S. I., and Kulikowski, C. 1991. *Computer Systems That Learn: Classification and Prediction Methods from Statistics, Neural Networks, Machine Learning, and Expert Systems*. San Francisco, Calif.: Morgan Kaufmann.

Woo, Ji Young, Sung Min Bae, and Sang Chan Park. "Visualization method for customer targeting using customer map." *Expert Systems with Applications* 28.4 (2005): 763-772.